ORIGINAL ARTICLE

DIFFICULTY INDEX, DISCRIMINATION INDEX, SENSITIVITY AND SPECIFICITY OF SINGLE BEST ANSWER QUESTIONS TO ASSESS MEDICAL STUDENTS' PERFORMANCE IN SEND UP EXAMINATION AT LAHORE MEDICAL & DENTAL COLLEGE

Uzma Zafar, Zaima Ali, Attiqa Khalid, Saima Zaki*, Faiqa Jabeen Naeem, Muhammad Asad Chaudhry

Department of Physiology, Lahore Medical & Dental College, *Department of Obstetrics and Gynaecology, Jinnah Hospital, Lahore, Pakistan

Background: Item analysis of the single best answer question based assessment is a simple and valuable method that can help the examiners to check the reliability and validity of this examination tool. This study aimed to determine difficulty index and discrimination index of single best answer question (SBAQ) based send-up Physiology paper of 1st Prof MBBS, and to check sensitivity and specificity of SBAQs to predict undergraduate medical students' performance. **Methods:** This was an observational study. A retrospective review of undergraduates' examination performance was done after approval from Institutional Review Board. Theory paper consisted of 45 SBAQs. Each SBAQ comprised a stem followed by five options. One option was correct and rest four were distractors. Data was analysed on SPSS-22. Difficulty and discriminative index was calculated. Receiver operating characteristics (ROC) curve was generated. The sensitivity and specificity of SBAQ was determined. **Results:** Out of 45 items, 33 (73%) were having difficulty index within the acceptable range (30–70%), 3 (6.7%) were very easy with the difficulty index of more than 70%, and 9 (20%) very difficult with the index of less than 30%. Mean±SD difficulty index was 47.45±17.83. Discrimination index revealed 33 items within acceptable range, 10 were poor and 2 items revealed negative scoring. Mean±SD discrimination index was 0.30±0.19. ROC curve revealed AUC of 0.86 (p=0.000; CI= 0.803-0.90) with sensitivity of 84% and specificity of 89%. **Conclusion:** Mean difficulty index of send-up paper was within acceptable limits. However, discrimination index identified 12 items that need revision.

Keywords: Difficulty index, Discrimination index, Single best answer question, Sensitivity, Specificity

INTRODUCTION

Assessment —a key component of any educational system, is more than a tool to measure performance of the students. Competency assessment is essential in medical education as it can modify the learning approach of the students and improve their performance.[1,2] Properly designed assessment can increase the learning drive of students and check the cognitive, psychomotor and affective domains.[3] Assessment provides an effective feedback to students as well as facilitators. Students can improve their learning strategies to achieve learning objectives while it helps facilitators to identify the gaps between the teaching methods and learning objectives.[4] Inappropriately designed assessments can result in undesirable results of competencies.[5]

Recently, huge effort has been allocated by different medical education programs and authorities at under- and postgraduate levels to enhance authenticity and effectiveness of assessments.[1] Many methods have been designed to measure different aspects of competencies. Any single assessment method cannot be considered perfect as each format has its pros and cons. Assessment design should be reliable, valid, acceptable, feasible, and have educational impact to be authentic.[6] Assessments can be designed to check the four components: knows, knows how, shows how, and does.[3] Single Best Answer Questions (SBAQs) have become a popular assessment design in many educational domains including medical education.[7] This method has an edge over other assessment designs as extensive part of the course can be assessed in a short time frame. Efficiently designed SBAQs can effectively identify strengths and shortcomings of students as well as discriminate between the best and weak. It can guide the facilitators on their educational strategies.[8] Properly designed SBAQs can be used to check interpretation and application of knowledge (high cognitive level) rather than simple recall of facts, although it can be challenging and demands a lot of effort by examiner.[9]

Item analysis of the SBA questions is simple and valuable method to help examiners to check the reliability and validity of this examination tool.[10] It can inform examiners about how difficult the SBAQ was (difficulty index) and its ability to discriminate between weak and strong students (discrimination index). Flawed multiple choice questions can affect the performance of high achievers as well as borderline students.[11]

This study aimed to determine difficulty index, discrimination index, sensitivity and specificity of the individual SBAQs of send-up Physiology paper of 1st

Prof MBBS for predicting our undergraduate medical students' performance.

MATERIAL AND METHODS

This was an observational study conducted at Department of Physiology, Lahore Medical and Dental College, Lahore from Nov 2021 to Feb 2022 after approval from Institutional Review Board (Ref No. LM&DC/1702830). A retrospective review of examination performance of 1st year MBBS students of session 2021 appearing in the send-up examination of Physiology theory was done.

The paper comprised of 45 SBAQs with one mark to each SBAQ (item) and no negative marking. Each SBAQ comprised of a stem and 5 options. One option was the correct option (key) and the rest 4 were incorrect (distracters). Pass percentage, difficulty index, and discrimination index of each SBAQ were calculated.

Difficulty index (also termed easiness index) is the proportion of students who attempted the item correctly. Its value ranges between 0 to 100%. Acceptable range for difficulty index is 30–70%. Items having values less than 30% or greater than 70% are considered as difficult/easy questions respectively.[12–15]

$$\text{Difficulty index} = (CN/TN) \times 100$$

where CN is number of students who correctly attempted the item and TN is total number of students who attempted the item.

Discrimination index is a measure to discriminate between the high and low achievers. The range for discrimination index is between -1.00 to +1.00. The expected response is that the high achievers should pick the correct response more frequently compared to the lower achievers. Items having discrimination index of ≥0.4 are considered as good items, 0.3–0.39 are reasonable, 0.2–0.29 are marginal or borderline items, ≤0.19 are poor and these items should be revised or removed from the assessment.[12–15]

$$\text{Discrimination index} = [(CH/NH)-(CL/NL)]$$

where CH and CL stand for number of correct responses in high and low achiever group, and NH and NL stand for number of students in high and low achiever group. Top and bottom ⅓ were taken as high and low achievers.

Data was analysed on SPSS-22. Difficulty and discriminative index was calculated. Receiver operating characteristics (ROC) curve was generated and sensitivity and specificity of SBAQ were determined. Area under curve (AUC) was calculated. AUC value of more than 0.7 was considered as an acceptable value for prediction and a value of >0.8 was considered as very good value. Correlation between discrimination and difficulty index was checked with the Pearson test.

RESULTS

A total of 147 candidates of 1st professional MBBS attempted the SBAQ paper in the subject of Physiology.

Out of 147 candidates, 72 (49%) passed and 75 (51%) did not pass. There were 45 items in the question paper; 33 (73%) items were having difficulty index within the acceptable range of 30–70%, 3 (6.7%) were very easy with the difficulty index of >70% and 9 (20%) items were very difficult with the index of <30%. Difficulty index of the SBAQ paper was 47.45±17.83. Discrimination index revealed 33 items within acceptable range, 10 were poor and 2 items revealed negative scoring and were defective items (Tables 1–3).

Discrimination index was 0.30±0.19. Receiver operating characteristics (ROC) curve analysis revealed AUC of 0.86 (p=0.000, CI=0.803–0.90) with sensitivity of 84% and specificity of 89% (Figure-1). There was no correlation between difficulty and discrimination index with Pearson test (r=0.155, p=0.350).

Table-1: Overall results of SBAQs (n=147)

SBAQs	Frequency	Percentage
Pass	72	49.0
Fail	75	51.0

Table-2: Difficulty index of SBAQs (n=45)

Difficulty index	SBAQ [n (%)]	Interpretation
30–70%	33 (73.3)	Acceptable
>70%	3 (6.7)	Very easy
<30%	9 (20)	Very difficult

Table-3: Discrimination index of SBAQs (n=45)

Discrimination index	SBAQ [n (%)]	Interpretation
>0.35	21 (46.7)	Excellent
0.25 to 0.35	9 (20)	Good
0.2 to 0.24	3 (6.7)	Acceptable
0-0.2	10 (22.2)	Poor
Negative value	2 (4.4)	Defective item

Table-4: Difficulty index and Discrimination index of each item of SBAQs

SBAQ	Difficulty index	Discrimination index	SBAQ	Difficulty index	Discrimination index
1	34.7	0.29	24	47.6	0.53
2	91.8	0.08	25	81	0.20
3	45.6	0.1	26	35.4	0.31
4	38.1	0.63	27	51.7	0.37
5	28.6	0.04	28	8.8	0.10
6	33.3	0.36	29	32.7	0.51
7	62.6	0.37	30	54.4	0.08
8	76.2	-0.1	31	54.3	0.46
9	29.9	0.35	32	59.2	0.28
10	47.6	0.51	33	66.7	0.40
11	23.8	0.14	34	63.9	0.21
12	48.3	0.25	35	45.6	0.37
13	59.2	0.53	36	68	0.6
14	39.5	0.48	37	31.3	0.10
15	45.6	0.48	38	50.3	0.43
16	56.5	0.01	39	42.9	0.15
17	48.3	0.64	40	25.9	0.27
18	51.7	0.77	41	26.5	0.11
19	64.6	0.46	42	68.7	0.35
20	64.6	0.10	43	29.9	0.29
21	68.7	0.35	44	38.10	0.27
22	53.7	0.41	45	23.10	0.06
23	16.3	-0.09			
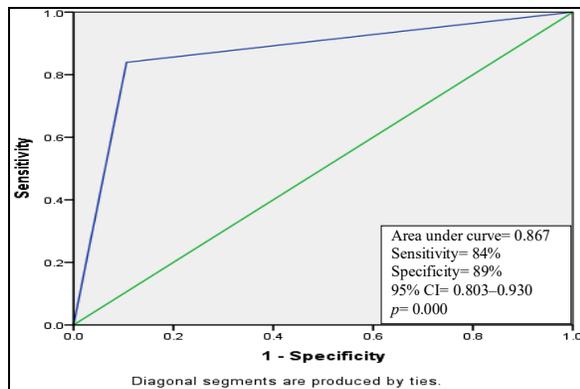
Figure-1: Receiver operating characteristic curve analysis to predict sensitivity and specificity of SBAQs

DISCUSSION

In this study, difficulty index and discrimination index of SBAQs of the send paper of Physiology was determined. These indices were employed to post validate the items individually and also to assess whether the items were properly constructed or not. Currently in medical curriculum there is a rising trend to use multiple choice questions or SBAQs to assess knowledge. The individual items of SBAQs can check memory based core knowledge of the candidates and if constructed thoughtfully can also assess higher cognitive domains such as application and problem solving skills.[16–17] It is very essential to control the quality of assessment.

In the present study, 33 (73%) items were having difficulty index within the optimum range, however 9 (20%) items were very difficult and 3 (6.7%) were very easy. The easy items in this study were placed at the start of the theory paper as warm-up items and they were based upon core knowledge which is essential for the students. Therefore they were not removed. The difficult items were checked for incorrect key, controversial areas or any language flaws. The purpose of adding difficult item by the assessor was to identify the top scorers.[12–15] Higher the value of difficulty index, more easy the item is; and higher the value of discrimination index better it can discriminate between the knowledgeable and non-knowledgeable candidates.[18] There is reciprocal relationship between difficulty and discrimination index. However, when Pearson test was applied no significant correlation was observed between the difficulty and discrimination indices. Previous studies based upon item analysis in the subjects of Physiology and Pathology from India reported positive correlation between difficulty and discrimination index.[19,20]

On Discrimination index analysis, 33 (73%) items were having discrimination index within or above the acceptable range and 12 items out of 45 needed revision. There were 10 items (out of 12) having discrimination index within the poor range (0–0.2) and in the remaining 2 items it was negative. Items having discrimination index between 0–0.2 were checked for the content, cognition level and the answer key. Eight out of the 10 items were primarily based upon the core knowledge. They were formulated to check recall and understanding of the candidates. Therefore, poor discrimination index between 0–0.2 was not considered a problem in these items, whereas 2 out of the 12 items where discrimination index was less than zero (negative value) were checked for the language error, typographical flaws, face validity and the answer key. As no major error was found, it was assumed that learning objectives were not properly conveyed to the students in the teaching session of the particular topics and students might have misunderstood. These points were planned to be clarified in the paper discussion session with students.

In this study sensitivity and specificity of SBAQs was determined by the ROC curve analysis and it was observed that SBAQs were 84% specific and 89% sensitive in predicting passed and failed candidates. This result is in concordance with another study based on psychometric analysis at Department of Medical Sciences, University Sains Malaysia, reporting discriminative ability of multiple choice question of more than 0.8 by ROC curve.[12] A study from India also reported multiple choice questions as an effective tool to assess students' performance.[19]

SBAQs are employed as an assessment tool for entry tests and also for under- and postgraduate pragmatic examinations.[21] They are used because of objectivity, reproducibility, decreased element of observer bias and comprehensive coverage of the subject in less time.[22,23] A well-formulated SBAQ based assessment can measure cognition level based upon Bloom's Taxonomy and also differentiate between high and low scorers.[24] While formulating an assessment a reliable and valid reference pool is required having multiple choice questions of known difficulty and discrimination index.[25] However, a cross-sectional study in 20 medical schools of UK concluded that there was pseudo-impression of competency with SBAQ based assessment. They found very short answer questions well correlated with students' performance than the SBAQs.[21]

One of the limitations of this study is that it is conducted in one institutional setting only. Therefore generalizability of the results to other educational institutions or programs is reserved.

CONCLUSION & RECOMMENDATIONS

Mean difficulty index of the send-up paper was within the acceptable limits. However, discrimination index identified 12 items that need revision. Discrimination index and difficulty index are useful tools to post-validate the assessment. In order to control quality of the future assessments, SBAQs with known difficulty and discrimination indices should be selected.

REFERENCES

1. Epstein RM. Assessment in medical education. New Eng J Med 2007;356(4):387–96.
2. Karim K, Edwards R, Dogra N, Anderson I, Davies T, Lindsay J, *et al.* A survey of the teaching and assessment of undergraduate psychiatry in the medical schools of the United Kingdom and Ireland. Med Teach 2009;31(11):1024–9.
3. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990;65(9 Suppl):S63–7.
4. Bienstock JL, Katz NT, Cox SM, Hueppchen N, Erickson S, Puscheck EE. To the point: Medical education reviews—providing feedback. Am J Obstet Gynecol 2007;196(6):508–13.
5. Gulikers JT, Bastiaens TJ, Kirschner PA. A five-dimensional framework for authentic assessment. Educ Technol Res Dev 2004;52(3):67–86.
6. van der Vleuten C. Validity of final examinations in undergraduate medical training. BMJ 2000;321(7270):1217–9.
7. Kalms E. Creating assessments as an active learning strategy: too time-consuming for benefit gained? Med Educ Online 2019;24(1):1649572.
8. Tan L, McAleer J. The introduction of single best answer questions as a test of knowledge in the final examination for the fellowship of the Royal College of Radiologists in Clinical Oncology. Clin Oncol (R Coll Radiol) 2008;20(8):571–6.
9. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc 2012;62(2):142–7.
10. Patel RM. Use of Item analysis to improve quality of multiple choice questions in II MBBS. J Educ Technol Health Sci 2017;4(1):22–9.
11. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ 2008;42(2):198–206.
12. Taib F, Yusoff MSB. Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. J Taibah Univ Med Sci 2014;9(2):110–4.
13. Shaikh S, Kannan SK, Naqvi ZA, Pasha Z, Ahamad M. The role of faculty development in improving the quality of multiple-choice questions in dental education. J Dent Educ 2020;84(3):316–22.
14. Mahjabeen W, Alam S, Hassan U, Zafar T, Butt R, Konain S, *et al.* Difficulty index, discrimination index and distractor efficiency in multiple choice questions. Ann Pak Inst Med Sci 2017;13(4):310–5.
15. Gajjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. Indian J Community Med 2014;39(1):17–20.
16. Al-Wardy NM. Assessment methods in undergraduate medical education. Sultan Qaboos Univ Med J 2010;10(2):203–9.
17. Touissi Y, Hjiej G, Hajjioui A, Ibrahimi A, Fourtassi M. Does developing multiple-choice questions improve medical students' learning? A systematic review. Med Educ Online 2022;27(1):2005505.
18. Kocdar S, Karadag N, Sahin MD. Analysis of the difficulty and discrimination indices of multiple-choice questions according to cognitive levels in an open and distance learning context. Turkish Online J Educ Technol 2016;15(4):16–24.
19. Rao C, Prasad HK, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. Int J Educ Psychol Res 2016;2(4):201–4.
20. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discrimination indices of MCQs in formative exam in physiology. South-East Asian J Med Educ 2013;7(1):45–50.
21. Sam AH, Westacott R, Gurnell M, Wilson R, Meeran K, Brown C. Comparing single-best-answer and very-short-answer questions for the assessment of applied medical knowledge in 20 UK medical schools: Cross-sectional study. BMJ Open 2019;9(9):e032550.
22. Fatima S, Ahmed ST, Hashmat S, Iftikhar A, Hashmi TA. The relationship of difficulty and discrimination indices with physiology teaching and learning. Pak J Med Dent 2021;10(2):109–13.
23. Al-Rukban MO. Guidelines for the construction of multiple choice questions tests. J Family Community Med 2006;13(3):125–33.
24. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. BMC Med Educ 2007;7:49.
25. Kadam AA, Verma A, Khare AS. Are difficulty index and discrimination index useful tools for assessing the quality of an MCQ? —A cross sectional study. Eur J Mol Clin Med 2021;8(4):1672–6.

Address for Correspondence:

Dr Uzma Zafar, Department of Physiology, Lahore Medical & Dental College, Lahore, Pakistan. **Cell:** +92-300-4058682
Email: uzma.zargham@gmail.com

Contribution of Authors:
UZ: Concept, data collection, analysis, write-up, proof reading
ZA: Concept, data collection, analysis, write-up, proof reading
AK: Idea, data collection, write-up, analysis
SZ: Data interpretation, write-up
FJN: Data collection, analysis, write-up
MAC: Data collection and interpretation with analysis

Conflict of Interest: None to declare
Disclaimer: None
Funding disclosure: None